

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications from the Center for Plant
Science Innovation

Plant Science Innovation, Center for

2019

FAD2 Gene Radiation and Positive Selection Contributed to Polyacetylene Metabolism Evolution in Campanulids1[OPEN]

Tao Feng

Ya Yang


Lucas Busta

Edgar B. Cahoon

Hengchang Wang

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/plantscifacpub>

 Part of the [Plant Biology Commons](#), [Plant Breeding and Genetics Commons](#), and the [Plant Pathology Commons](#)

This Article is brought to you for free and open access by the Plant Science Innovation, Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Center for Plant Science Innovation by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Tao Feng, Ya Yang, Lucas Busta, Edgar B. Cahoon, Hengchang Wang, and Shiyou Lü

FAD2 Gene Radiation and Positive Selection Contributed to Polyacetylene Metabolism Evolution in Campanulids¹[OPEN]

Tao Feng,^a Ya Yang,^b Lucas Busta,^c Edgar B. Cahoon,^c Hengchang Wang,^{a,2} and Shiyou Lü^{d,2,3}

^aCAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

^bDepartment of Plant and Microbial Biology, University of Minnesota, Twin Cities, St. Paul, Minnesota 55108

^cCenter for Plant Science Innovation and Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, Nebraska, 68588

^dState Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan 434200, China

ORCID IDs: 0000-0002-0489-2021 (T.F.); 0000-0002-0102-9986 (L.B.); 0000-0002-7277-1176 (E.B.C.); 0000-0003-0449-2471 (S.L.).

Polyacetylenes (PAs) are bioactive, specialized plant defense compounds produced by some species in the eudicot clade campanulids. Early steps of PA biosynthesis are catalyzed by Fatty Acid Desaturase 2 (*FAD2*). Canonical *FAD2*s catalyze desaturation, but divergent forms can catalyze hydroxylation, conjugation, acetylation, and epoxygenation. These alternate reactions give rise to valuable unusual fatty acids, including the precursors to PAs. The extreme functional diversity of *FAD2* enzymes and the origin of PA biosynthesis are poorly understood from an evolutionary perspective. We focus here on the evolution of the *FAD2* gene family. We uncovered a core eudicot-wide gene duplication event giving rise to two lineages: *FAD2-α* and *FAD2-β*. Independent neofunctionalizations in both lineages have resulted in functionally diverse *FAD2-LIKEs* involved in unusual fatty acid biosynthesis. We found significantly accelerated rates of molecular evolution in *FAD2-LIKEs* and use this metric to provide a list of uncharacterized candidates for further exploration of *FAD2* functional diversity. *FAD2-α* has expanded extensively in Asterales and Apiales, two main clades of campanulids, by ancient gene duplications. Here, we detected positive selection in both Asterales and Apiales lineages, which may have enabled the evolution of PA metabolism in campanulids. Together, these findings also imply that yet uncharacterized *FAD2-α* copies are involved in later steps of PA biosynthesis. This work establishes a robust phylogenetic framework in which to interpret functional data and to direct future research into the origin and evolution of PA metabolism.

Plant metabolism, known to produce large amounts of diverse molecules, provides striking examples of biological complexity that have intrigued biologists for decades (Pichersky and Gang, 2000; Weng et al., 2012).

¹This work was supported by the National Natural Science Foundation of China (NSFC) (grant no. 31570186), a grant from the National Science Foundation (Plant Genome IOS-13-39385), a post-doctoral research fellowship from National Science Foundation (Post-doctoral Research Fellowship in Biology IOS-1812037), and the Major Program of the National Natural Science Foundation of China (NSFC) (31590823).

²Senior authors.

³Author for contact: shiyoulu@hubu.edu.cn.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Shiyou Lü (shiyoulu@hubu.edu.cn).

T.F. collected the data and performed analyses; T.F., Y.Y., and L.B. wrote the article with contributions of all the authors; H.W. and S.L. conceived the original screening and research plans; all authors read and approved the final version of this manuscript.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.19.00800

When compared with relatively highly conserved plant primary metabolism, plant specialized metabolism is more diverse, and generates numerous lineage-specific specialized metabolite classes (Milo and Last, 2012), such as glucosinolates in Brassicales (Halkier and Gershenzon, 2006), betalains in Caryophyllales (Brockington et al., 2015), and acylsugars in Solanaceae (Moghe et al., 2017). Many specialized metabolites are essential in responding to biotic stress (e.g. herbivory defense), attracting pollinators, and communicating with other species (e.g. allelopathy; Arimura and Maffei, 2017). Specialized metabolism is often found in a lineage-specific pattern (Moghe and Kruse, 2018), which enables using comparative approaches to aid biosynthetic pathway discovery and the study of pathway evolution in a phylogenetic framework. With the recent advances in functional and comparative genetics, new approaches integrating functional and genomic studies in the context of phylogenetic frameworks have made fundamental advances in understanding evolution of plant specialized metabolites (e.g. Brockington et al., 2015; Moghe et al., 2017; Lopez-Nieves et al., 2018). Previous

studies exploring the occurrence of lineage-specific metabolites have revealed that gene duplication followed by functional diversification is of central importance to this process. An example of this is the duplication followed by neofunctionalization of both *DODA* and *CYP76AD* genes in the origin of betalain biosynthesis (Brockington et al., 2015).

Plant polyacetylenes (PAs) are a class of compounds characterized by multiple alkynyl groups in their carbon skeleton (Bu'Lock and Smith, 1967; Boll and Hansen, 1987). PAs exhibit a wide range of bioactivities including allergenic, antibacterial, and antifungal activities and are also known as natural pesticides (Minto and Blacklock, 2008). PAs have been detected in 24 plant families, but seem to occur in a substantial number of species in only seven plant families (Bu'Lock and Smith, 1967; Negri, 2015). The Asterales family Asteraceae contains the most structurally diverse PAs, with more than 1100 out of about 2000 PAs that have been identified in plants so far (Konovalov, 2014). A second hotspot of PA diversity is found in the Apiales families Apiaceae and Araliaceae, which contain most of the C17-PAs, for example, falcarinol and falcarindiol (Bohlmann et al., 1974; Hansen and Boll, 1986). In addition, PAs also widely occur in species of Pittosporaceae (Apiales) and Campanulaceae (Asterales). Both of these families, together with Asteraceae, Apiales, and Araliaceae, all belong to the eudicot clade campanulids, making campanulids the center of diversity for PAs (Fig. 1). Outside campanulids, PAs have also been detected sporadically in a few species in, for example, Olacaceae and Santalaceae (Negri, 2015).

Although PAs are extremely diverse in their chemical structures, the core acetylenic molecules are derived from unsaturated fatty acids, which are intermediates in plant lipid biosynthesis (Fig. 1). In plants, three biosynthetic pathways for PAs have been proposed (Minto and Blacklock, 2008; Negri, 2015): the crepenynate pathway, the steariolate pathway, and the tarirate pathway. Each of these pathways starts with acetylation of different intermediates of the fatty acid metabolism (Fig. 1). Little is known about the steariolate and tarirate pathways, and the downstream PAs are only sporadically found in distantly related species (Negri, 2015). In this study, we focus on the better characterized crepenynate pathway, which in plants has only been found in campanulids, and gives rise to the majority of the known PAs (Fig. 1). Early intermediates of this pathway have been identified by radiochemical studies (Barley et al., 1988; Knispel et al., 2013), and the pathway has been examined in diverse campanulid lineages (Lee et al., 1998; Cahoon et al., 2003; Carlsson et al., 2004; Nam and Kappock, 2007; Busta et al., 2018). Linoleic acid is converted to crepenynic acid by installing the first acetylenic bond in the Δ^{12} position (the 12th carbon relative to the fatty acid head group; this notation is used consistently throughout). The enzyme (Δ^{12} -acetylenase) catalyzing the first committed step in the crepenynate pathway was characterized in the Asteraceae plant *Crepis alpina*

and was found to be similar to Fatty Acid Desaturase 2 (*FAD2*; Lee et al., 1998). Additional Δ^{12} -acetylenases have subsequently been isolated and characterized from other PA-producing plants in campanulids, such as parsley (*Petroselinum crispum*), English ivy (*Hedera helix*), and carrot (*Daucus carota*; Cahoon et al., 2003; Busta et al., 2018). All these enzymes were found to be encoded by *FAD2* family members, which were traditionally known as Δ^{12} -desaturases (Higashi and Murata, 1993; Okuley et al., 1994b), indicating that these acetylenases are *FAD2*s with diverged function. Further stepwise desaturation and acetylenation at Δ^{14} and Δ^{16} positions generate the core acetylenic molecules, each of which is further modified in a lineage-specific manner, giving rise to diverse PAs across campanulids (Christensen and Brandt, 2006; Minto and Blacklock, 2008; Konovalov, 2014). Recently, Busta et al. (2018) identified two Δ^{14} -desaturases (catalyzing the second step of crepenynate pathway) in carrot, which are also encoded by *FAD2* genes. These studies imply that *FAD2* plays important roles in the initiation of PA biosynthesis in all lineages investigated so far.

FAD2 is a member of the fatty acids desaturases family, which introduce double bonds into monounsaturated fatty acids, producing polyunsaturated fatty acids (Harwood, 1980; Sperling et al., 2003). The *FAD2* gene was first characterized in *Arabidopsis thaliana*, which has a single copy (Okuley et al., 1994b). Subsequent studies identified multiple copies from a variety of oil crops, for example, canola (*Brassica napus*; Jung et al., 2011), olive (*Olea europaea*; Hernández et al., 2005), soybean (*Glycine max*; Damude et al., 2006), sunflower (*Helianthus annuus*; Martínez-Rivas et al., 2001), and cotton (*Gossypium hirsutum*; Zhang et al., 2009). These studies indicated that *FAD2* specifically acts on the Δ^{12} position of the oleic acid, introducing a second double bond into the chain, thus producing the polyunsaturated linoleic acid. However, a number of *FAD2-LIKEs* have been characterized from distantly related plants that exhibit diverse functions, for example, hydroxylation (van de Loo et al., 1995; Broun and Somerville, 1997), conjugation (Cahoon et al., 1999, 2001; Hornung et al., 2002), and epoxidation and acetylenation (Lee et al., 1998; Cahoon et al., 2003). In contrast with the canonical *FAD2s* that perform Δ^{12} desaturation reactions, the *FAD2-LIKEs* are versatile in their substrate specificities, regioselectivities, and catalytic activities, resulting in the formation of diverse unusual fatty acids, including crepenynic acid, the precursor to PAs. Plant oils rich in unusual fatty acids are of high value in human nutrition and/or in industrial utilization (Biermann et al., 2011), and there are increasing interests to genetically engineer the synthesis of unusual fatty acids in crops (Napier, 2007). Studies using domain swapping, site-directed mutagenesis, and in vitro directed evolution have progressed in identifying key residues that exert control over reaction partitioning between the "unusual" catalytic activities and canonical desaturase activity (Broun et al., 1998; Lee et al., 1998; Rawat et al., 2012). Taken together,

these studies suggest that plant *FAD2* genes have undergone extensive diversification and are involved in many plant specialized metabolic processes, including the biosynthesis of plant defense-related PAs. Despite these interesting and important aspects of *FAD2*s, there lacks a robust phylogenetic framework in which to interpret existing functional data from diverse lineages.

In this study we reconstruct the phylogeny of the *FAD2* gene family across green plants to explore the origin and evolution of unusual fatty acids synthesis, with emphasis on crepenynic acid, a precursor to many PAs. We sought to understand the extreme functional diversity of *FAD2* from an evolutionary perspective. We provide a robust phylogenetic framework to facilitate

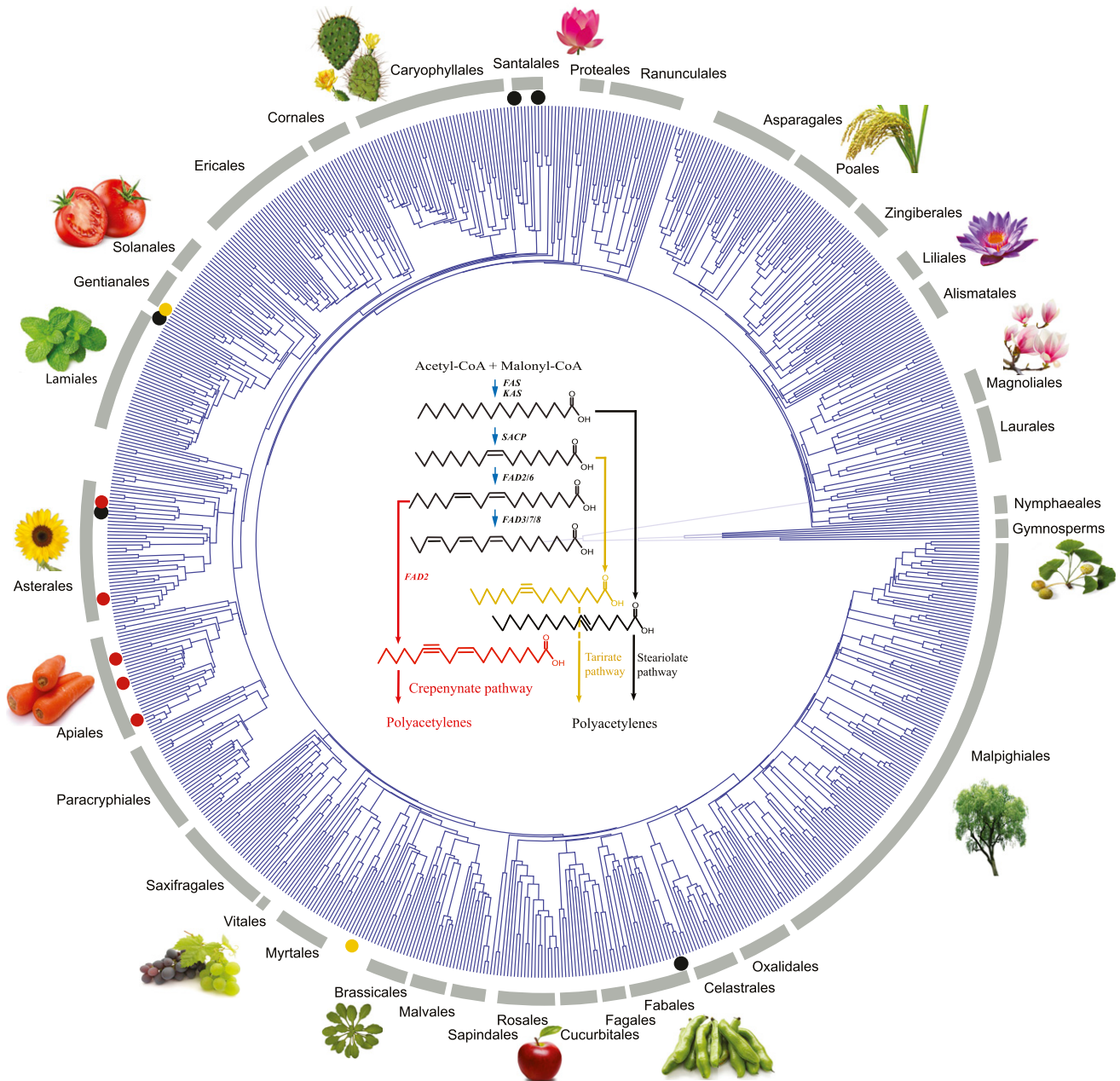


Figure 1. The proposed biosynthesis pathways of PAs and the distribution of PAs in seed plants. Enzyme abbreviations: *FAD*, fat acid desaturase; *FAS*, fatty acid synthase; *KAS*, 3-ketoacyl-ACP synthase; *SACP*, stearoyl-ACP desaturase. The distribution of the crepenynate pathway, steariolate pathway, and tartrate pathways are indicated by red, black, and brown circles, respectively. The distribution data are summarized from literature and also PlantFADb (<https://plantfad.org/>), Crepenynate: Apiaceae, Araliaceae, Pittosporaceae, Asteraceae and Campanulaceae; Steariolate: Asteraceae, Lamiaceae, Olacaceae, Opiliaceae, Santalaceae, and Fabaceae; Tartrate: Picramniaceae and Lamiaceae. The tree topology is modified from Magallón et al. (2015), which included 87% families of angiosperm and was based on five plastid and nuclear markers.

the future deciphering the remaining steps of the crepenynate-derived PA biosynthesis pathway, and to understand the genetic mechanisms underlying the origin and evolution of PA metabolism in campanulids.

RESULTS AND DISCUSSION

The Evolutionary History of *FAD2* in Green Plants

BLAST search and phylogenetic analysis revealed that *FAD2* is present in all major plant lineages, including green algae, indicating an origin predating the common ancestor of green plants. Phylogenetic reconstruction produced a gene tree that was concordant with the relationships among major lineages of green plants (Bowe et al., 2000; Puttick et al., 2018). For example, gymnosperms, ferns, and mosses + liverworts were resolved as paraphyletic and successively sister to flowering plants (Fig. 2; Supplemental Fig. S1). The liverworts and mosses were sister to each other with a bootstrap percentage (BS) of 79, which corresponds to the recently defined "Setaphyta" clade (Puttick et al., 2018). Topology within the angiosperms was congruent with accepted phylogeny with *Amborella* resolved as the

earliest diverging lineage and *Aquilegia* + *Papaver* placed as sister to all other eudicots (The Angiosperm Phylogeny Group, 2016). Within the angiosperms, the evolution of *FAD2* was characterized by a gene duplication event at the base of core eudicots (Fig. 2). Following the divergence of Ranunculales (represented by *Aquilegia* and *Papaver* in this study), a gene duplication event yielded two subclades, here termed as *FAD2- α* (BS = 65) and *FAD2- β* (BS = 63), respectively (Fig. 2). This duplication event is likely associated with the Arabidopsis Gamma whole genome duplication event (At- γ) that occurred at ~135 Ma, shared by all core eudicots (Bowers et al., 2003; Tang et al., 2008). Most notably, *FAD2- α* experienced multiple gene duplication events in campanulids (Fig. 2, gray circles), resulting in a large number of *FAD2* homologs in Asterales and Apiales.

Neofunctionalization Events Occurred Independently in Eudicot *FAD2* Lineages

FAD2 is the key enzyme responsible for biosynthesis of polyunsaturated fatty acids by desaturating the monounsaturated fatty acids (Okuley et al., 1994a).

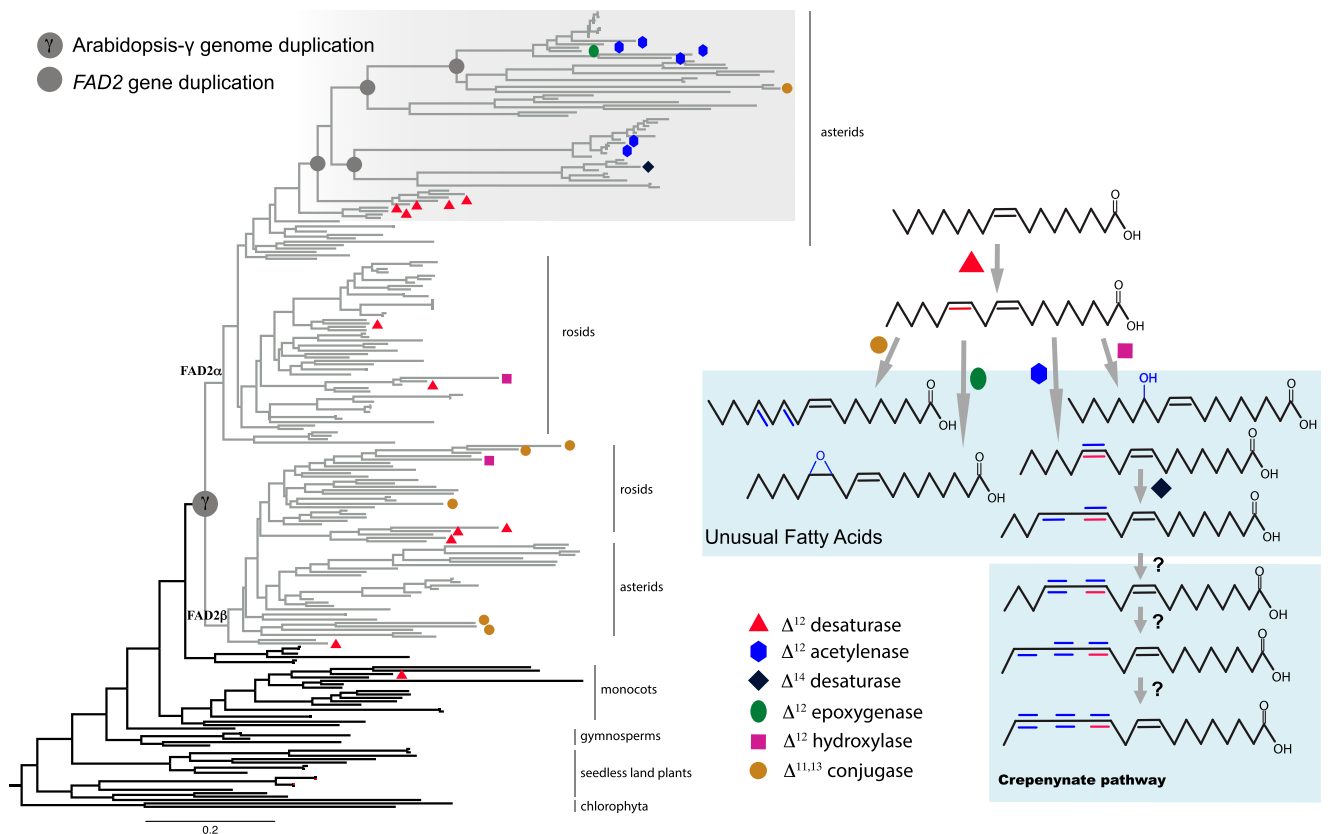


Figure 2. Maximum likelihood (ML) gene tree of *FAD2* sequences in green plants. Core eudicot homologs are indicated in gray. The gene duplication event shared by all sampled eudicots and campanulids-specific gene duplication events are indicated on the tree as gray circles. Previously characterized enzymes are labeled with different symbols after tips, and their catalytic activities are shown in the diagram of fatty acid biosynthesis pathway on the right.

Divergent forms of *FAD2* associated with activities other than desaturation, such as conjugation, hydroxylation, acetylation, and epoxygenation, have been identified across eudicots in, for example, Brassicaceae, Euphorbiaceae, Cucurbitaceae, and Punicoidae (Supplemental Table S1). These *FAD2-LIKE* enzymes redirect metabolic flux from primary fatty acid modification and ubiquitous lipid biosynthesis into specialized biosynthesis pathways, generating diverse biologically important molecules. Previous phylogenetic analyses, conducted with only a few characterized genes to infer relationships between *FAD2* and *FAD2-LIKEs*, were unable to resolve the phylogenetic origins of the latter (Cahoon et al., 2003; Iwabuchi et al., 2003; Cao et al., 2013; Lakhssassi et al., 2017). This was partially because *FAD2-LIKEs* were always subtended by longer branches than canonical *FAD2* members, making *FAD2-LIKEs* cluster together in neighbor-joining analyses, likely due to long branch attraction. With a much denser sampling across green plants, our maximum likelihood phylogenetic analysis revealed that *FAD2-LIKEs* are polyphyletic and the encoded enzymes have arisen independently many times in eudicots from ancestral desaturases (Fig. 2).

Accelerated Molecular Evolution Rate in *FAD2-LIKEs*

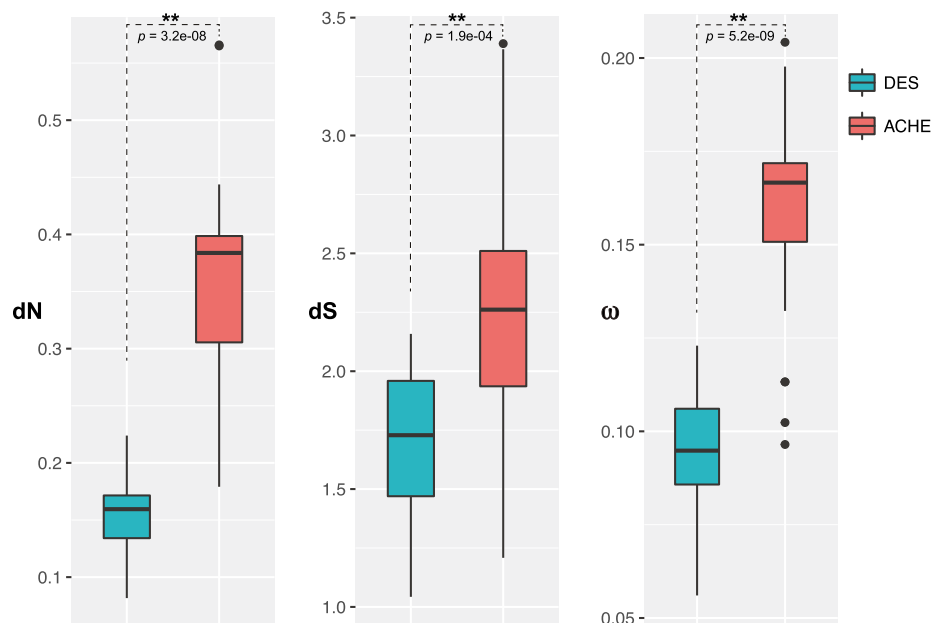
Visual examination of the green plant *FAD2* gene tree reconstructed from amino acids (Fig. 2) revealed that the branches associated with *FAD2-LIKEs* are substantially longer than those associated with the *FAD2*. To test the hypothesis that nonsynonymous substitutions were significantly higher in *FAD2-LIKEs* compared with canonical *FAD2*, we calculated the nonsynonymous substitutions (dN), synonymous substitutions (dS), and their ratio (ω) using the free

ratio model as implemented in codeml. We defined dS or dN for a gene as the total substitutions from the eudicot most recent common ancestor node to the terminal tip of that gene. We examined branches leading to the 43 genes (Supplemental Table S1) that had been experimentally characterized and assigned to either *FAD2* or *FAD2-LIKEs*. The dN values of *FAD2* genes ranged from 0.08 to 0.22 (mean 0.16), whereas that of *FAD2-LIKEs* ranged from 0.18 to 0.57 (mean 0.36). The dS values for *FAD2-LIKEs* (median \pm SD 2.31 ± 0.32) were generally higher than that for *FAD2* (1.68 ± 0.92). Welch's *t* test indicated that the difference of dN and dS between *FAD2* and *FAD2-LIKEs* were significant (Fig. 3; Supplemental Table S2), with $P = 3.2e-08$ and $1.9e-04$ for dN and dS, respectively.

Despite the caveats that the tested branches are partially shared and are therefore not completely independent, and mutations observed at terminal branches may not be fixed substitutions, our results suggest accelerated rates of molecular evolution in *FAD2-LIKEs*. The ratios of dN/dS (ω) were also significantly higher among the *FAD2-LIKEs* (0.16 ± 0.05) than that among *FAD2* (0.09 ± 0.08 ; Fig. 3). Together, these results suggest multiple shifts to significantly higher rates of nonsynonymous substitutions in *FAD2-LIKEs* than in *FAD2*, and that such can be used as a molecular signature for detecting potential *FAD2-LIKEs* that are yet uncharacterized.

FAD2-LIKEs produce a variety of fatty acids that together are called "unusual fatty acids" and are industrially valuable (Diedrich and Henschel, 1991; Napier, 2007). The hydroxylated fatty acid ricinoleic acid is a precursor for chemical conversion to many industrial products, such as emulsifiers, inks, and lubricants (Jaworski and Cahoon, 2003). More than 450 different unusual fatty acids have been detected in plants (Ohlrogge et al., 2018); however, the biosynthetic

Figure 3. Boxplots of sequence variation. Shown are differences in dN, dS, and ω , among *FAD2* (green) and *FAD2-LIKEs* (red). Significance of the comparisons were determined by Welch two-sample *t*-test. The original estimated parameters and details of the statistical tests are shown in Supplemental Table S2 and Supplemental Table S3.



pathways of most of them are yet unknown. Previous functional studies have identified members of *FAD2* gene family as important players for their biosynthesis [for review, see Napier (2007)]. To facilitate the identification of enzymes underlying the biosynthesis of the fatty acids, we provided a list of *FAD2* genes (Supplemental Table S4) showing accelerated molecular evolution rates based on dN/dS estimations. These genes will likely be good to include in future functional characterization efforts aimed at discovering new *FAD2-LIKE* genes. Integrating such functional analyses in the robust phylogenetic framework described here will significantly broaden our perspective on the functional diversity of *FAD2* enzymes.

FAD2- α Extensively Expanded in Campanulids

Initial BLAST search and phylogenetic analysis using high-quality genomes identified elevated numbers of *FAD2* homologs in several plant species, namely sunflower (Asterales, 29 copies), ginseng (*Panax ginseng*, Apiales, 52 copies), and carrot (Apiales, 24 copies), as compared with the 1–5 copies in other plant species. To trace the gene family expansion with a denser taxon sampling, we reconstructed the evolutionary history of *FAD2* using three genomes (sunflower, carrot, and ginseng) and 27 transcriptomes across the campanulids. The resulted gene tree suggests that *FAD2* expansion is not limited to sunflower, carrot, and ginseng,

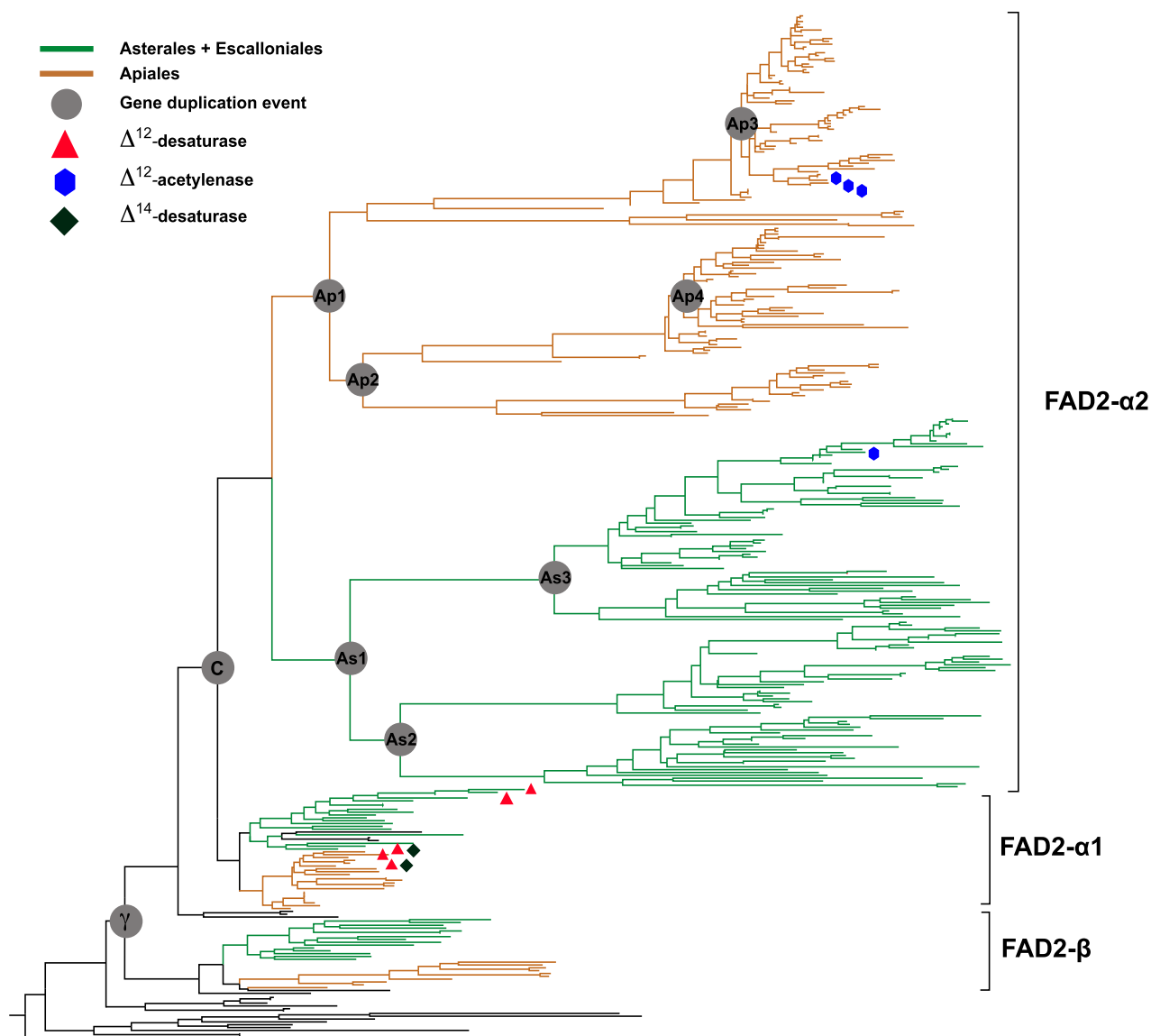


Figure 4. ML gene tree of *FAD2* sequences from campanulids. Asterales-Escalloniales lineages are in green, and Apiales lineages in brown. Gene duplication events are indicated on the tree as gray circles with γ indicating the Arabidopsis gamma genome duplication, C indicating the campanulids-specific gene duplication, As representing Asterales-specific gene duplication, and Ap representing the Apiales-specific gene duplication events.

but is widespread in Asterales, Apiales, and Escalloniales (3 out of 7 orders in campanulids). In addition to the FAD2- α and FAD2- β clades shared by core eudicots, following the divergence of Aquifoliales (Supplemental Fig. S2), FAD2- α splits into two subclades, here termed as FAD2- α 1 and FAD2- α 2, respectively (Fig. 4). Under FAD2- α 2, four and three subclades of Asterales-Escalloniales and Apiales homologs are further recovered, respectively (Fig. 4; Supplemental Fig. S2). At the species level, most representatives examined in this study harbor multiple highly similar homologs that are likely resulted from tandem duplication (Supplemental Fig. S2). This assumption is supported by the localization of the *FAD2* homologs in carrot genomes (Busta et al., 2018). These duplications have resulted in 20–60 copies in the species of Asterales, Escalloniales, and Apiales, whereas only a few copies (1–5) are typically found in other lineages.

Comparison of the gene tree with the phylogeny of campanulids (Tank and Donoghue, 2010; Magallón et al., 2015; Stull et al., 2018) indicates a total of eight gene duplication events at the family level or above in campanulids. These include a duplication event that occurred in the common ancestor of core campanulids (Fig. 5). In addition, at least two and three gene

duplication events occurred in the common ancestor of Apiales and the common ancestor of Asterales-Escalloniales, respectively (Fig. 5; As1-3, Ap1-2), and two more happened before the divergence of Apiaceae and Araliaceae (Fig. 5; Ap3-4).

Ancestral Sequence Reconstruction Shows Shifts of Functionally Important Residues in FAD2- α 2 Proteins Subsequent to Gene Duplication

Previous functional studies using site-directed mutagenesis have identified five residues that are important for the function of *FAD2* enzymes (Cahoon and Kinney, 2005; Meesapyodsuk et al., 2007; Rawat et al., 2012). To test whether there are shifts in these five residues following gene duplication, we carried out ancestral sequence reconstruction using both amino acid sequences and coding sequences. Subsequent to the campanulids-specific gene duplication C, a shift of M³²² to I³²² was observed in FAD2- α 2 clade (Fig. 6). M³²² was suggested to be the diagnostic residues for canonical Δ^{12} -desaturase (Cahoon and Kinney, 2005), and its replacement by I³²² may enable noncanonical functions evolved. Further substitutions of A¹⁰² by G¹⁰² and V¹⁵⁴

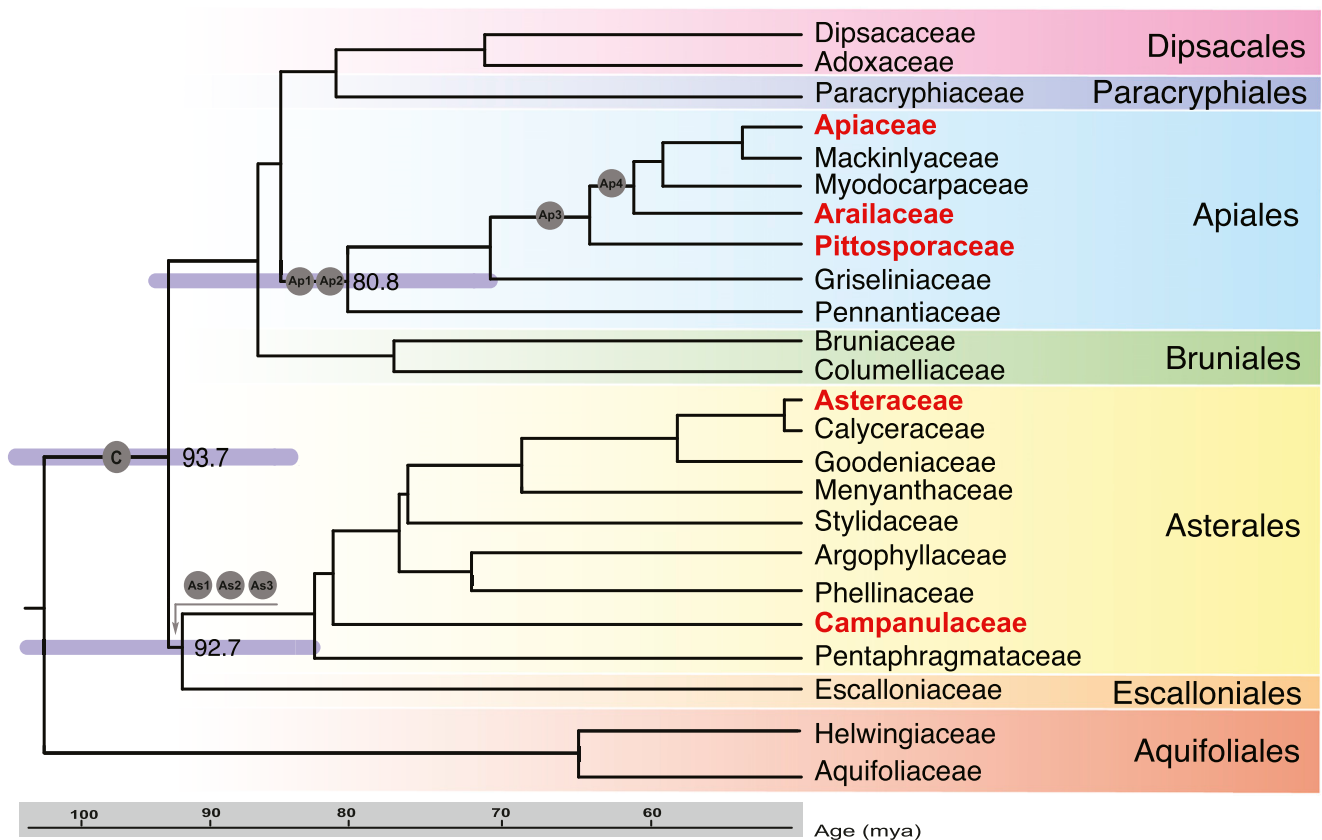


Figure 5. The phylogeny of campanulids. Gene duplication events are indicated with gray circles. The duplication events labeled here correspond to the ones labeled in Figure 4, and their locations in the phylogeny of campanulids are based on their phylogenetic placements in the gene tree. Polyacetylene-producing lineages are in red, and nonpolyacetylene lineages are in black. The tree topology and diversification dating are based on Magallón et al. (2015) and Stull et al. (2018).

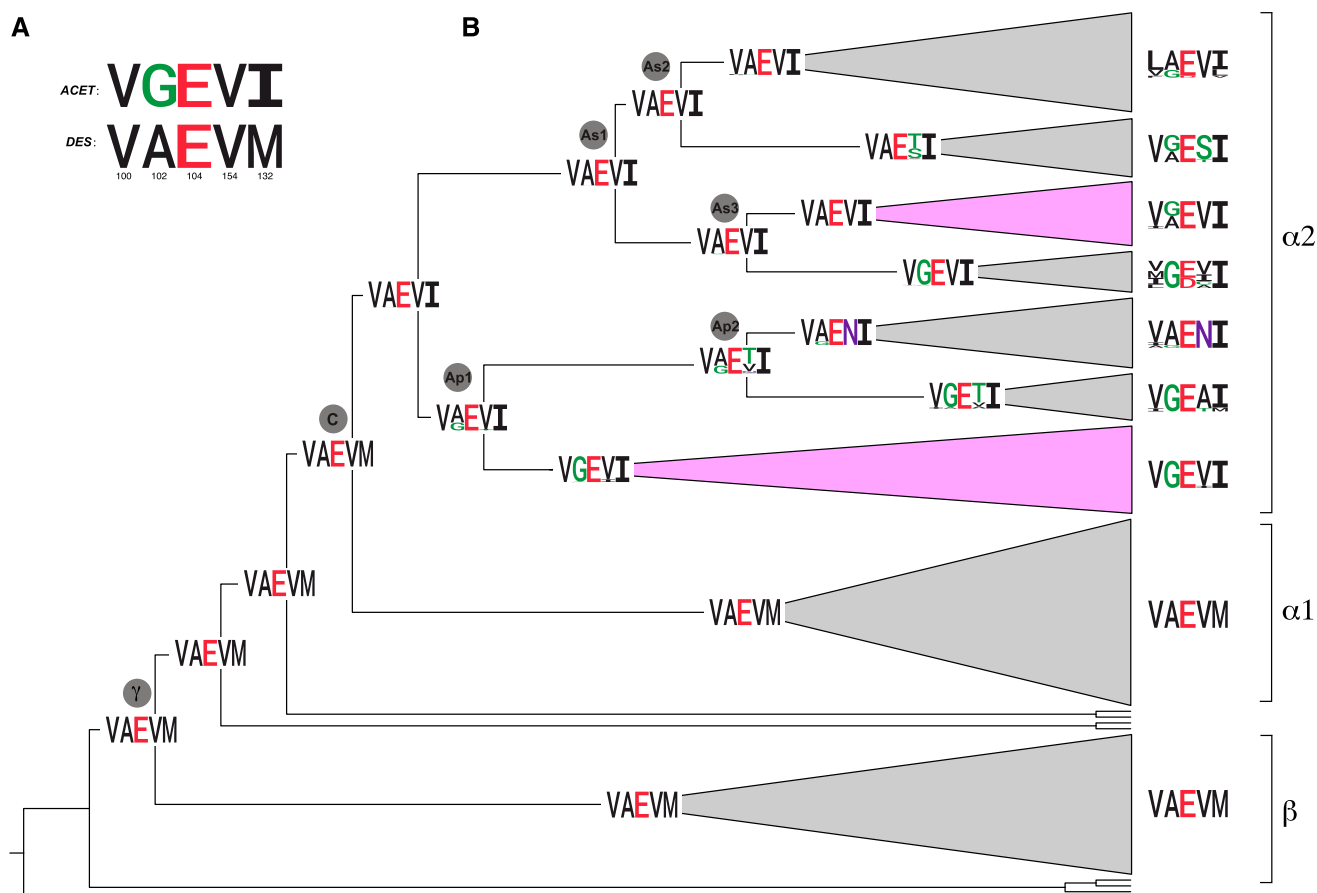


Figure 6. Reconstruction of five functionally important amino acids on the campanulids *FAD2* phylogeny show shifts in key residues in the *FAD2*- $\alpha 2$ lineages subsequent to gene duplications. A, Proportional logo plots (percent contribution of each base at each position) of the five residues identified across functionally characterized canonical *FAD2* (DES: desaturase) and divergent *FAD2* (ACET: acetylenase) homologs included in this study. B, Ancestral sequence reconstructions of amino acids by empirical Bayesian inference on a maximum likelihood phylogeny of a reduced campanulids *FAD2* sequence alignment. Logo plots at nodes indicate the posterior probability of an amino acid at each of the five sites. Lineages that are known to be involved in polyacetylene biosynthesis are labeled in pink.

by N¹⁵⁴, S¹⁵⁴, or A¹⁵⁴ were observed during the evolution of *FAD2*- $\alpha 2$, whereas both residues were conserved in *FAD2*- $\alpha 2$ and *FAD2*- β (Fig. 6). The residue in position 154 defines the regioselectivity of the *Claviceps purpurea* *FAD2* desaturase, which is primarily a Δ^{12} -desaturase and has traces of Δ^{15} -desaturase activity. A single mutation of Val to Ala at this position significantly increases the Δ^{15} -desaturase activity from 0.37% to 11.57% (Meesapyodsuk et al., 2007). It is likely that the substitution of V¹⁵⁴ by N¹⁵⁴, S¹⁵⁴, or A¹⁵⁴ observed in *FAD2*- $\alpha 2$ would lead to diversification of the regioselectivity of *FAD2*- $\alpha 2$ proteins. Taken together, these observations suggest that the *FAD2*- $\alpha 2$ clade includes many members with divergent function.

In fact, all members of *FAD2*- $\alpha 2$ characterized so far encode enzymes that catalyze reactions other than desaturation (Supplemental Table S1). Several previously characterized genes fall in the *FAD2*- $\alpha 2$ clade (Fig. 6), and they perform Δ^{12} -epoxidation (Lee et al., 1998; Hatanaka et al., 2004) or Δ^{12} -acetylation reactions

(Lee et al., 1998; Cahoon et al., 2003; Busta et al., 2018), producing epoxy fatty acids and crepenynic acid, respectively. Epoxy fatty acids are known as defensive substances in vivo as they can strongly inhibit blast fungus (Greene and Hammock, 1999; Moran et al., 2000). Crepenynic acid is the first step in the PA-generating crepenynate pathway, and the derived PAs are known as natural pesticides (Boll and Hansen, 1987; Minto and Blacklock, 2008). Together, this evidence suggests that *FAD2*- $\alpha 2$ genes acquired new functions in defense against herbivores and pathogens. Further studies testing the function of *FAD2*- $\alpha 2$ across campanulids families will provide insights into the functional diversity of *FAD2*- $\alpha 2$.

Variation in Selective Pressures among Branches and Positive Selection in *FAD2*- $\alpha 2$

Given the inferred neofunctionalizations, especially in *FAD2*- $\alpha 2$, that follow gene duplication events, next

we tested the hypothesis that selective pressures varied among branches leading to major clades in *FAD2*. We chose sunflower and carrot as the representative of Asterales and Apiales, respectively, because of the availability of reference genomes and that several *FAD2* genes from both species have been functionally characterized (Supplemental Table S1). In addition, both species, together with other members of Asteraceae and Apiaceae, are the main sources of naturally occurring PAs. We performed a series of nested hypotheses tests (Table 1) to compare the selective pressures (ω) among branches leading to major clades of *FAD2* (Supplemental Fig. S3) using the branch model in software package PAML v4.9 (Phylogenetic Analysis by Maximum Likelihood; Yang, 2007). The results indicated that the free-ratio model, H5, fits the data the best (Table 1). This model allows the values of dN/dS to vary among all the tested branches. The second-best hypothesis was H4, which suggested shifts in selective pressure subsequent to gene duplications (Table 1). This shift of selective pressures resulted in the internal branches leading to *FAD2-β*, *FAD2-α1*, *FAD2-α2*, *FAD2-Asα2*, and *FAD2-Apα2* having different dN/dS values. Likelihood ratio test were significant for H5-H4 ($P < 0.01$) and H4-H3 ($P < 0.01$). The other hypotheses, H1, H2, and H3, which also assume different ω values between *FAD2-β* and *FAD2-α*, were all significantly better than one-ratio null hypothesis H0 (Table 1). Comparisons between all alternative hypothesis (H1, H2, H3, H4, and H5) are also significant (Table 1). Under all alternative hypotheses that allowed dN/dS to vary across branches, we observed the highest ω values in *FAD2-α2* subsequent to gene duplication events (Table 1).

To test whether there was positive selection acting on *FAD2* genes, we next implemented the adapted branch-site model aBSREL (adaptive branch-site random effects likelihood; Smith et al., 2015). The full adaptive hypothesis ($\omega_k^b > 1$) identified that 15 branches out of 112 were under positive selection (Fig. 7). All these 15 branches were in the *FAD2-α2* clade with four in *FAD2-Ap-α2* and 11 in *FAD2-β* (Fig. 7). It is notable that the ancestral branches b2, b5/b7 leading to the acetylenases that catalyze the first step of PA biosynthesis were under positive selection (Fig. 7). Other previously characterized genes (encoding conjugase, epoxygenase) also experienced episodic positive selection during their evolutionary history (b6, b8, b9 in Fig. 7). No positive selection was detected in the *FAD2-β* or *FAD2-α1* clade, together with the observation that the ω values for them were relatively low (Table 1; Supplemental Fig. S3), indicating that these lineages were under purifying selection. Consistent with this assumption, genes in *FAD2-β* and *FAD2-α1* maintained the ancestral desaturation function (Supplemental Table S1).

Implications of the Phylogeny for the Origin and Evolution of PA Metabolism

We have identified a *FAD2* subclade in campanulids, *FAD2-α2*, which has substitutions in key residues

Table 1. Parameter estimates in PAML under nested branch models using phylogeny of *FAD2* homologs from sunflower, carrot, and an additional 16 functionally characterized campanulid *FAD2*s

Hypothesis	Model ^a	Background ω		Foreground ω^b		lnL	AIC ^c	P
		ω_0	ω_1	ω_2	ω_3			
H0	M0 (1)	$\omega_0 = 0.0955$		n/a		-34444.3357	68890.6713	
H1	M2 (2)	$\omega_0 = 0.1134$		$\omega_{1-7} = 0.0926$		-34441.3610	68886.7220	0.0147*
H2	M2 (3)	$\omega_0 = 0.1134$		$\omega_{1,3} = 0.0631$; $\omega_{2,4,5,6,7} = 0.1016$		-34424.2389	68854.4779	<0.01*, 5e-09*
H3	M2 (6)	$\omega_0 = 0.1117$		$\omega_1 = \infty$; $\omega_3 = 0.0593$; $\omega_2 = 1.1451$; $\omega_{4,5} = 0.1617$; $\omega_{6,7} = 0.0457$		-34268.4925	68548.9851	<0.01*, <0.01*, <0.01*
H4	M2 (8)	$\omega_0 = 0.1111$		$\omega_1 = \infty$; $\omega_3 = 0.0591$; $\omega_2 = 0.1752$; $\omega_4 = 1.1671$; $\omega_5 = 0.1616$; $\omega_6 = \infty$; $\omega_7 = 0.0467$		-34258.7784	68533.5567	<0.01*, <0.01*, <0.01*, 7e-06*
H5	M1 (112)			ω ranges from 0.004 to 1.233		-34027.9770	68279.9540	all <0.01*

^aBranch models implemented in PAML4, followed by the number of parameters used by the model. ^bBranch classes 1–7 refer to the branches labeled in Supplemental Figure S3; infinity of ω value (∞) indicates lack of synonymous substitution (dS). ^cAkaike Information Criteria (AIC = $-2\ln L + 2N$) is used to rank models.

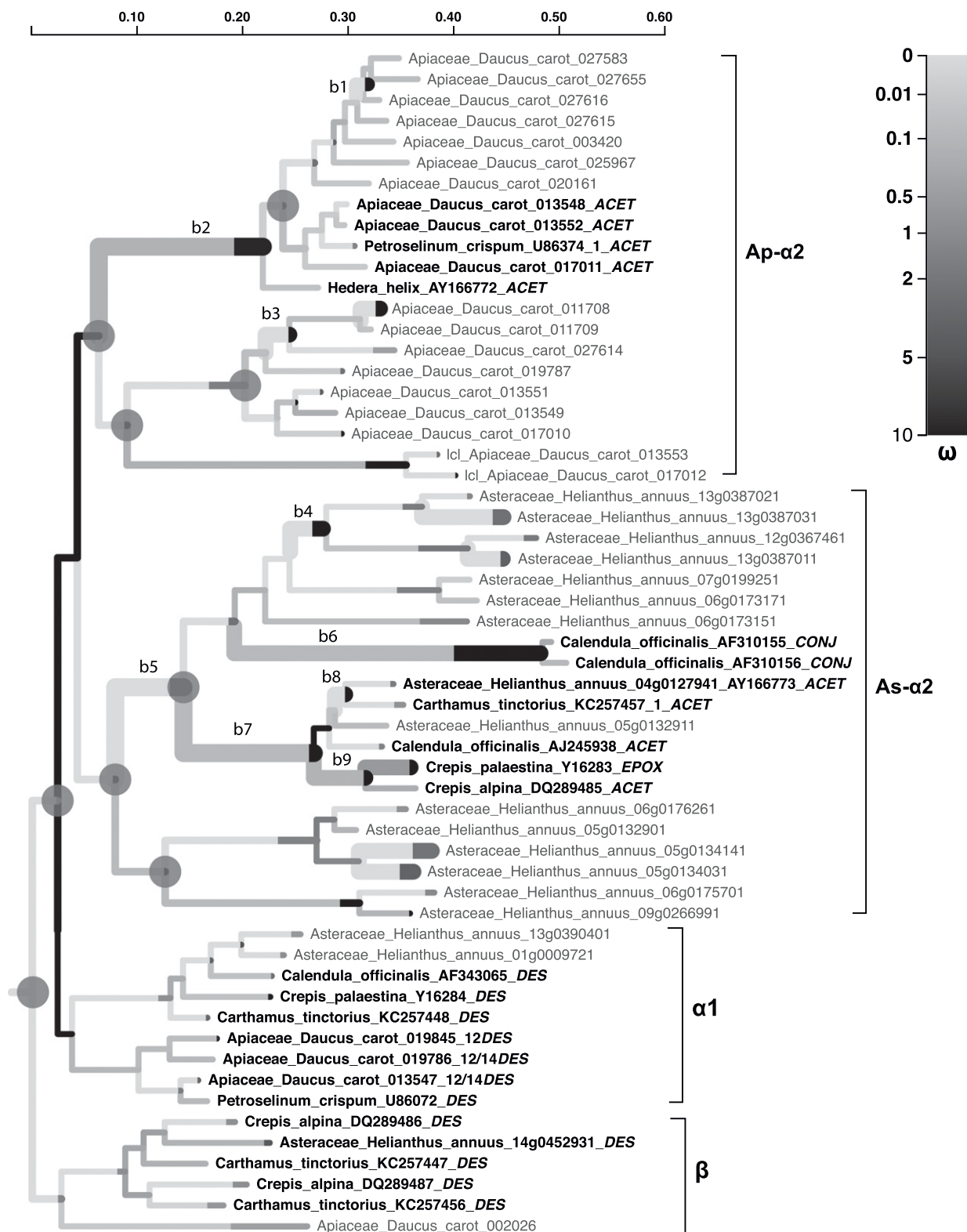


Figure 7. Selection analyses of the campanulids *FAD2* with aBSREL models. The branch is annotated according to the inferred distribution of ω ; the branch is partitioned according to the proportion of sites in a particular class (f_k^b); and the color of the segment depicts the magnitude of the corresponding ω_k^b . Branches that are identified as having experienced positive selection are labeled as thicker lines ($P < 0.05$ with correction for multiple testing). Gene duplication events are indicated with gray circles. Characterized genes are labeled as bold text. *ACET*, acetylenase; *CONJ*, conjugase; *DES*, desaturase; *EPOX*, epoxynase.

Table 2. Parameter estimates of branches on which positive selection are detected under the adaptive aBSREL model compared with the null MG94×REV model.Significance of the comparison ($P < 0.05$) was determined by likelihood ratio test (LRT). Branch labels refer to Figure 7.

Branch	LRT	P	ω Distribution Over Sites (%)
b1	11.4731	0.0011	$\omega_1 = 0.00$ (99.03); $\omega_2 = 214$ (0.97)
b2	20.4477	0.0000	$\omega_1 = 0.135$ (82); $\omega_2 = 8.65$ (18)
b3	15.0688	0.0002	$\omega_1 = 0.00$ (95.8); $\omega_2 = 3090$ (4.2)
b4	5.8009	0.0197	$\omega_1 = 0.00$ (87); $\omega_2 = 2690$ (13)
b5	4.3683	0.0411	$\omega_1 = 0.00$ (80); $\omega_2 = 4.53$ (20)
b6	4.7515	0.0337	$\omega_1 = 0.167$ (86); $\omega_2 = 24.3$ (14)
b7	5.0482	0.0290	$\omega_1 = 0.0806$ (96.4); $\omega_2 = 302$ (3.8)
b8	11.8177	0.0009	$\omega_1 = 0.00$ (97.3); $\omega_2 = 166$ (2.7)
b9	5.1270	0.0278	$\omega_1 = 0.133$ (96.6); $\omega_2 = 10.8$ (3.4)
Daucus_carot_011708	6.2793	0.0154	$\omega_1 = 0.00$ (99.03); $\omega_2 = 14.2$ (0.97)
Helianthus_annuus_05g0134141	4.5004	0.0384	$\omega_1 = 0.00$ (74); $\omega_2 = 2.25$ (26)
Helianthus_annuus_05g0134031	6.4991	0.0138	$\omega_1 = 0.00$ (75); $\omega_2 = 3.55$ (25)
Helianthus_annuus_13g0387011	4.5101	0.0382	$\omega_1 = 0.00$ (94); $\omega_2 = 3.86$ (6.3)
Crepis_palaestina_Y16283	6.9848	0.0108	$\omega_1 = 0.562$ (97.6); $\omega_2 = 27.5$ (2.4)
Helianthus_annuus_13g0387031	5.8543	0.0192	$\omega_1 = 0.00$ (87); $\omega_2 = 2.47$ (13)

contrasting to canonical *FAD2* (desaturase) and has undergone episodes of positive selection. The distinct characteristics of the *FAD2- α 2* clade and its evolutionary diversification have not been recognized previously and its role in producing valuable unusual fatty acids remains to be explored.

Previous functional studies (Supplemental Table S1) showed that *FAD2* genes encode enzymes that are diversified in catalytic activity (desaturation, hydroxylation, and acetylation) as well as regioselectivity (Δ^{12} , Δ^{14} position of the fatty acid carbon chain). Several *FAD2* homologs from Asteraceae and Apiaceae have been shown involved in the first catalytic step of PA biosynthesis (Lee et al., 1998; Cahoon et al., 2003; Busta et al., 2018). Our phylogenetic analyses indicate that there are divergent *FAD2*s with unknown functions, particularly within the *FAD2- α 2* lineage. It seems likely that these uncharacterized *FAD2- α 2* lineages may be involved in the further oxidation reactions in the crepenynate pathway, namely the stepwise desaturation and acetylation at Δ^{14} and Δ^{16} positions (Fig. 2). This in turn suggests that many, if not all, of the central enzymatic steps in the crepenynate pathway may have arisen via duplication and neofunctionalization of *FAD2* (the *FAD2- α 2* lineage in particular) in campanulids. Such cumulative processes, by which enzymes catalyzing earlier steps of a pathway emerge first, and duplication of the gene encoding the first enzyme creates enzymes catalyzing later reactions, have been observed for core fatty acid biosynthesis (Li-Beisson et al., 2013). Functional characterization of the homologs in the *FAD2- α 2* clade from plant species that produce PAs, for example sunflower and carrot, will help to determine whether this same mechanism prevails in the evolution of the PA biosynthesis pathway.

The distribution of PAs and patterns of *FAD2* duplication in the campanulids (Fig. 5) raises questions about whether PA biosynthesis arose independently in Asterales and Apiales. Presumably, the Asterales and Apiales-specific duplications would have facilitated

the evolution of acetylenase activity, the first step in the crepenynate pathway, which would support the notion of independent evolution of PA biosynthesis in these two lineages. Interestingly, Dipsacales did not undergo further *FAD2* duplication except for the core campanulids shared duplication C (Fig. 5), meaning that this lineage should be a target of future PA analyses to shed light on whether acetylenase activity may have arisen after the divergence of Apiales and Dipsacales (Fig. 5). Overall, the phylogenetic framework provided here will be an excellent resource with which to further explore questions surrounding independent evolution. In any event, it is interesting to consider what evolutionary advantages may have led to the invention of PAs in campanulids. Fatty acid-derived PAs, such as falcarinol and falcarindiol, have been identified as antifungal molecules, which can inhibit germination of spore in different fungi (Bu'Lock and Smith, 1967). Consistent with the antipathogen properties of PAs, expression of the related genes is strongly induced by fungal elicitation (Kirsch et al., 2000; Cahoon et al., 2003; Busta et al., 2018). Our analyses detected positive selections in multiple *FAD2- α 2* lineages, including the acetylenases that are known to be involved in PA biosynthesis (Fig. 7). Together, these observations imply that plant-pathogen interaction may have stimulated the evolution of PAs in Asterales and Apiales. Further studies to test the functions of *FAD2- α 2* genes across the phylogeny of campanulids would help to further clarify the evolution of PA biosynthesis pathways.

CONCLUSION

We provide a comprehensive phylogeny for *FAD2* genes across green plants. We also characterize in detail the molecular evolution of *FAD2* within campanulids. By establishing the comparative framework, we (1) reveal a complex evolutionary history of *FAD2* genes, which is characterized by gene duplication and

functional diversification; (2) describe an approach to distinguish canonical *FAD2* from functionally divergent *FAD2-LIKEs* using both the molecular evolution rate and the phylogenetic relationships to other *FAD2* genes; and (3) identify a unique *FAD2- α 2* clade in campanulids with implications for future functional studies aimed at characterizing further enzymes involved in PA biosynthesis. Ultimately, this improved understanding of diversity among *FAD2* genes will facilitate ongoing research into the biosynthesis of unusual fatty acids and fatty acid-derived natural products and the origin and evolution of PA metabolism, and may serve as a model for studies of tartrate and steariolate pathway evolution when those pathways are characterized in the future.

MATERIALS AND METHODS

Plant *FAD2* Sequences

Genome annotation data including both amino acid and coding sequences (CDS) were downloaded from public databases (Supplemental Table S5), including 46 flowering plants representing all major clades of APG IV (The Angiosperm Phylogeny Group, 2016), plus four gymnosperms, two monilophytes, one lycophytes, three bryophytes, and three green algae. All the amino acid and CDS were concatenated into a single FASTA file, respectively. A local BLAST database was generated using MAKEBLASTDB (-parse_seqids) implemented in the BLAST+ package v2.2.31 (Camacho et al., 2009) for the amino acid and coding sequence dataset, respectively. In addition, 44 previously characterized *FAD2* CDS were retrieved from literature (Supplemental Table S1). *Arabidopsis thaliana* *FAD2* protein sequence were used as query to search against the protein and nucleotide BLAST databases using BLASTp and tBLASTn with default settings, respectively. All positive hits were retrieved from the datasets using the BLASTDBCMD in BLAST+.

To infer the origins of the Δ^{12} -acetylenase, the first committing step of PA biosynthesis via the crepenynate pathway, 27 publicly available transcriptomes across the campanulids were included in this analysis (Supplemental Table S5), representing five of the seven orders of campanulids according to APG IV (The Angiosperm Phylogeny Group, 2016). At the moment no genome annotation or transcriptome is available for the two remaining campanulids orders Bruniales and Paracryphiales. The CDS from transcriptomes were concentrated into a single FASTA file, and the BLAST database was generated as described above. *Arabidopsis thaliana* *FAD2* amino acid sequence was used to search against the concentrated CDS dataset using SWIPE v2.1.0 (-p 3 -e 10; Rognes, 2011).

Alignment and Phylogeny Inference

Multiple-sequence alignment of the protein sequences was conducted in MAFFT v7.402 (Katoh and Standley, 2013) using default settings. As *FAD2* belongs to a large fatty acid desaturase gene family that share three Hxx(x)H motifs, sequences without the conserved motifs or were unalienable were manually removed from the initial alignment by visual examination. The remaining sequences were realigned for a second time with MAFFT and a tree was inferred using FastTree v2.1.5 (Price et al., 2010) with default settings. Previously characterized desaturases were used to identify gene families, and three well-defined clades corresponding to *FAD2*, *FAD3* (including *FAD7* and *FAD8*), and *FAD6* subfamilies were recovered and sequences of the *FAD2* family were extracted.

We then generated three datasets of *FAD2* for subsequent analysis: (1) *FAD2* gene lineage from 57 genomes across all green plants (from green algae to angiosperms), to reconstruct the broad-scale evolution history of *FAD2*; (2) *FAD2* gene lineage from three genomes and 27 transcriptomes across campanulids, to trace the origin of Δ^{12} -acetylenase; and (3) *FAD2* homologs from sunflower and carrot, representatives of Asterales and Apiales, respectively, to explore the signature of natural selection after gene duplication. Asterales and Apiales are the two main clades of campanulids, and the main source of natural occurring PAs.

To infer the phylogeny of *FAD2* genes for all three datasets, we conducted an iterative set of alignment and phylogenetic estimation steps. Initial alignments

were performed with MAFFT v7.407 (Katoh and Standley, 2013) using default settings and trimmed with Trimal v1.2 (-gt 0.05 -st 0.0001; Capella-Gutiérrez et al., 2009). Phylogenetic trees were estimated with FastTree v2.1.5 (Price et al., 2010), and the alignments were then refined with PASTA (-iter-limit = 3, aligner = mafft; Mirarab et al., 2015). The resulted alignments were again trimmed (trimal -gt 0.1), and trees were then inferred again using FastTree. Branches longer than 2 (absolute cutoff) or 10 times greater than their sister branch (relative cutoff) were most likely caused by distantly related homologs and/or transcriptome assembly artifacts and were removed using the script *trim_tips.py* (Yang et al., 2015). Sequences were realigned using MAFFT by using a variable scoring matrix (-ginsi,-allowshift-unalignlevel $\alpha_{max} = 0.8$) to minimize the overalign issues observed in previous MAFFT estimation (Katoh and Standley, 2016). Lower α_{max} values (0.5–0.7) were also used to generate various alignments and to examine the consistence among the derived phylogenies. The best-fit model was chosen according to the Bayesian Information Criterion (Schwarz, 1978) computed by modelFinder (Kalyaanamoorthy et al., 2017). Phylogenetic analyses were carried out by RAxML-NG v0.7.0b (Kozlov et al., 2018) with 20 parsimony starting trees, and 200 bootstrap replicates (-bs-metric tbe) were used to evaluate the variation in the dataset.

Rates of Sequence Divergence

To test whether there are significant differences in rates of molecular evolution between canonical versus divergent *FAD2* gene lineages, the number of dN per site, dS per site, and their ratio (ω , denoted by dN/dS) for each branch of the phylogeny derived from the green plant dataset were calculated under the free-ratios model (model = 1, Nsite = 0) using codeml (Yang, 2007). Branches with dS or dN > 3, suggesting saturation of substitutions, were not included in further analyses. The dS or dN for a gene is defined as total substitutions from the eudicot most recent common ancestor node to the terminal tip of that gene. The distance in terms of substitutions between the ancestor node and terminal tip was calculated using the biopython module *phylo* (Talevich et al., 2012). We categorized the 43 previously characterized genes (Table 2) into two groups based on their function: canonical *FAD2* (*DES*, short for desaturase) or divergent *FAD2* (*ACHE*, short for acetylenase, conjugase, hydroxylase, and epoxynase). The differences in dS, dN, and their ratio ω between *DES* and *ACHE* were evaluated using welch two-sample *t*-test in R.

Ancestral Sequence Reconstruction

To generate a numerically and computationally tractable dataset for ancestral sequence reconstruction, we subsample the campanulids *FAD2* gene tree (Supplemental Fig. S2) using python scripts (github.com/NatJWalker-Hale/DODA/tree/master/misc_scripts/python, accessed July 05, 2019). The strategy is to break the tree into paralogs, and then to sample within each paralog so that every genus had its longest sequence in the resulted alignment to minimize missing data. The final dataset has 153 CDS, includes all functionally characterized *FAD2* genes, and maintains within-paralogue diversity. A ML tree was inferred on amino acid sequences using the same procedure as described above. After confirming that the topology was consistent with the original tree (Supplemental Fig. S2), we performed marginal ancestral sequence reconstructions on this tree for amino acids and codons using IQ-TREE v1.6.8 (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017). Probabilities of ancestral states at each node were summarized using ggseqlogo implemented in R (Wagih, 2017; R Core Team, 2019). Five functionally important sites were visualized by summing over the probabilities of amino acids.

Estimation of the Variation in Selective Pressures (ω) among Branches

FAD2 homologs from genome annotations of sunflower and carrot, representatives of Asterales and Apiales, respectively, were used to explore the variation in selective pressures among the *FAD2* gene lineages. Five *FAD2* sequences from sunflower that had premature stop codons resulting from nonsense or frame shift mutation were removed. One *FAD2* homolog with a 96-bp deletion (encompassing the first His box) from carrot was also removed from the datasets. In addition, five and three characterized *FAD2* homologs from Asteraceae and Apiaceae, respectively, were included in the analysis. Tandem repeats sharing > 95% similarity in nucleotide sequence were removed to keep only one of them, resulting in 57 CDS in the final datasets. The protein

sequences were first aligned, and CDS alignment was generated by matching with the amino acid alignment using phyx (Brown et al., 2017). Unaligned codons were removed from the alignments, and topologies were estimated from the amino acid and CDS alignment using RAXML-NG. Confirming that the topologies are congruent, the amino acid trees were used in the codon model analyses.

A series of tests (Table 1) comparing the dN/dS (ω) for different clades or sets of branches (Supplemental Fig. S3) were performed with the branch model implemented in PAML v4.9 (Yang, 2007). The following nested hypotheses were formulated: H0, homogeneous selective pressure in all lineages leading to major *FAD2* clades (ω_0 , thick branches in Supplemental Fig. S3); H1, different selective pressure between stem branches leading to *FAD2*- α and - β ($\omega_0, \omega_1 = \omega_2-7$); H2, different selective pressure among stem branches leading to *FAD2*- α_1 , - α_2 , and - β ($\omega_0, \omega_1 = \omega_3, \omega_2 = \omega_4-7$); H3, based on H2 but allow shift in selective pressure after gene duplication events ($\omega_0, \omega_1, \omega_2, \omega_3, \omega_4 = \omega_5, \omega_6 = \omega_7$); H4, based on H3 but allowed ω to vary among numbered branches ($\omega_0, \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7$); H5, free-ratio model that allow all branches leading to major lineages of *FAD2* to vary (thick branches in Supplemental Fig. S3). Significance of the comparisons between hypothesis ($P < 0.05$) was determined by likelihood ratio test (LRT).

Selection Analyses

Next, branch-site models aBSREL (Smith et al., 2015) were used to detect episodic positive selection. The codon dataset and tree topology of campanulids (represented by sunflower, carrot) used in the branch model test above were used here to detect positive selection. We first set all branches as the foreground branch and ω distribution of each branch was inferred under null model MG94 ($k = 1; \omega_k \leq 1$) and the alternative model aBSREL ($k \geq 1; \omega_k > 1$). We then set *FAD2*- α_2 clade as foreground branch and infer the ω along each branch following the same procedure. Significance of the comparison between null model and alternative model was determined by χ^2 mixture distribution (Smith et al., 2015), and p value was corrected by Bonferroni's correction for multiple testing. All calculations and significance test were performed using the HYPHY package v2.3.13 (Pond et al., 2005).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Maximum likelihood phylogeny of green plant *FAD2*.

Supplemental Figure S2. Maximum likelihood phylogeny of campanulids *FAD2*.

Supplemental Figure S3. Maximum likelihood phylogeny of *FAD2* homologs from sunflower and carrot, plus characterized *FAD2* from close related species.

Supplemental Table S1. Functionally characterized *FAD2* genes used in the study.

Supplemental Table S2. Welch t test comparing dN, dS and ω between *DES* and *ACHE*.

Supplemental Table S3. Parameter estimation for characterized *FAD2* genes under free-ratio model using codeml. The genes are categorized as either *DES* or *ACHE*.

Supplemental Table S4. Uncharacterized *FAD2* genes that showed an accelerated molecular evolution rate ($\omega > 0.16$, the median of ω for *FAD2-LIKEs*).

Supplemental Table S5. Genome annotation data and transcriptomes used for BLAST search in this study.

ACKNOWLEDGMENTS

We thank Caifei Zhang for providing transcriptome data of Asteraceae species, and Samuel Brockington and Jinling Huang for commenting on an early version of the paper.

Received July 19, 2019; accepted August 3, 2019; published August 16, 2019.

LITERATURE CITED

- Arimura G, Maffei M** (2017) Plant Specialized Metabolism: Genomics, Biochemistry, and Biological Functions, 1st ed. CRC Press, Boca Raton
- Barley GC, Jones EHR, Thaller V** (1988) Crepenynate as a precursor of falcariol in carrot tissue culture. *In* J Lam, H Breteler, T Arnason, L Hansen, eds, Chem. Biol. Nat. Acetylenes Relat. Compd. Elsevier, Amsterdam, pp 85–91
- Biermann U, Bornscheuer U, Meier MAR, Metzger JO, Schäfer HJ** (2011) Oils and fats as renewable raw materials in chemistry. *Angew Chem Int Ed Engl* **50**: 3854–3871
- Bohlmann F, Burkhardt T, Zdero C** (1974) Naturally Occurring Acetylenes. Academic Press, London, New York
- Boll PM, Hansen L** (1987) On the presence of falcariol in araliaceae. *Phytochemistry* **26**: 2955–2956
- Bowe LM, Coat G, dePamphilis CW** (2000) Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc Natl Acad Sci USA* **97**: 4092–4097
- Bowers JE, Chapman BA, Rong J, Paterson AH** (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438
- Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GKS, Moore MJ, Smith SA** (2015) Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol* **207**: 1170–1180
- Broun P, Somerville C** (1997) Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic *Arabidopsis* plants that express a fatty acyl hydroxylase cDNA from castor bean. *Plant Physiol* **113**: 933–942
- Broun P, Shanklin J, Whittle E, Somerville C** (1998) Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science* **282**: 1315–1317
- Brown JW, Walker JF, Smith SA** (2017) Phyx: Phylogenetic tools for unix. *Bioinformatics* **33**: 1886–1888
- Bu'Lock JD, Smith GN** (1967) The origin of naturally-occurring acetylenes. *J Chem Soc C* 332–336
- Busta L, Yim WC, LaBrant EW, Wang P, Grimes L, Malyszka K, Cushman JC, Santos P, Kosma DK, Cahoon EB** (2018) Identification of genes encoding enzymes catalyzing the early steps of carrot polyacetylene biosynthesis. *Plant Physiol* **178**: 1507–1521
- Cahoon EB, Kinney AJ** (2005) The production of vegetable oils with novel properties: Using genomic tools to probe and manipulate plant fatty acid metabolism. *Eur J Lipid Sci Technol* **107**: 239–243
- Cahoon EB, Carlson TJ, Ripp KG, Schweiger BJ, Cook GA, Hall SE, Kinney AJ** (1999) Biosynthetic origin of conjugated double bonds: Production of fatty acid components of high-value drying oils in transgenic soybean embryos. *Proc Natl Acad Sci USA* **96**: 12935–12940
- Cahoon EB, Ripp KG, Hall SE, Kinney AJ** (2001) Formation of conjugated Δ^8, Δ^{10} -double bonds by Δ^{12} -oleic-acid desaturase-related enzymes: Biosynthetic origin of calendic acid. *J Biol Chem* **276**: 2637–2643
- Cahoon EB, Schnurr JA, Huffman EA, Minto RE** (2003) Fungal responsive fatty acid acetylases occur widely in evolutionarily distant plant families. *Plant J* **34**: 671–683
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL** (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* **10**: 421
- Cao S, Zhou X-R, Wood CC, Green AG, Singh SP, Liu L, Liu Q** (2013) A large and functionally diverse family of *Fad2* genes in safflower (*Carthamus tinctorius* L.). *BMC Plant Biol* **13**: 5
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T** (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973
- Carlsson AS, Thomaes S, Hamberg M, Stymne S** (2004) Properties of two multifunctional plant fatty acid acetylase/desaturase enzymes. *Eur J Biochem* **271**: 2991–2997
- Christensen LP, Brandt K** (2006) Bioactive polyacetylenes in food plants of the Apiaceae family: Occurrence, bioactivity and analysis. *J Pharm Biomed Anal* **41**: 683–693
- Damude HG, Zhang H, Farrall L, Ripp KG, Tomb J-F, Hollerbach D, Yadav NS** (2006) Identification of bifunctional Δ^{12}/ω^3 fatty acid desaturases for improving the ratio of ω^3 to ω^6 fatty acids in microbes and plants. *Proc Natl Acad Sci USA* **103**: 9446–9451

- Diedrich M, Henschel K-P (1991) The natural occurrence of unusual fatty acids Part 3: Acetylenic fatty acids. *Nahrung* **35**: 193–202
- Greene JF, Hammock BD (1999) Insulin-Like (IGF-I, IGF-II), Epidermal (EGF) and Alpha Transforming (TGF- α) growth factors in human breast cyst fluid (BCF). In KV Honn, LJ Marnett, S Nigam, EA Dennis, eds, *Eicosanoids and Other Bioactive Lipids in Cancer, Inflammation, and Radiation Injury*. Kluwer Academic/Plenum Publishers, New York, pp 471–477
- Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* **57**: 303–333
- Hansen L, Boll PM (1986) Polyacetylenes in Araliaceae: Their chemistry, biosynthesis and biological significance. *Phytochemistry* **25**: 285–293
- Harwood JL (1980) Plant acyl lipids: Structure, distribution, and analysis. In PK Stumpf and EE Conn, eds, *The Biochemistry of Plants: A Comprehensive Treatise*. Academic Press, New York, pp 1–55
- Hatanaka T, Shimizu R, Hildebrand D (2004) Expression of a *Stokesia laevis* epoxygenase gene. *Phytochemistry* **65**: 2189–2196
- Hernández ML, Mancha M, Martínez-Rivas JM (2005) Molecular cloning and characterization of genes encoding two microsomal oleate desaturases (*FAD2*) from olive. *Phytochemistry* **66**: 1417–1426
- Higashi S, Murata N (1993) An *in vivo* study of substrate specificities of acyl-lipid desaturases and acyltransferases in lipid synthesis in *Synchocystis* PCC6803. *Plant Physiol* **102**: 1275–1278
- Hornung E, Pernstich C, Feussner I (2002) Formation of conjugated Delta11Delta13-double bonds by Delta12-linoleic acid (1,4)-acyl-lipid-desaturase in pomegranate seeds. *Eur J Biochem* **269**: 4852–4859
- Iwabuchi M, Kohno-Murase J, Imamura J (2003) Delta 12-oleate desaturase-related enzymes associated with formation of conjugated trans-delta 11, cis-delta 13 double bonds. *J Biol Chem* **278**: 4603–4610
- Jaworski J, Cahoon EB (2003) Industrial oils from transgenic plants. *Curr Opin Plant Biol* **6**: 178–184
- Jung JH, Kim H, Go YS, Lee SB, Hur CG, Kim HU, Suh MC (2011) Identification of functional *BrFAD2-1* gene encoding microsomal delta-12 fatty acid desaturase from *Brassica rapa* and development of *Brassica napus* containing high oleic acid contents. *Plant Cell Rep* **30**: 1881–1892
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587–589
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* **30**: 772–780
- Katoh K, Standley DM (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**: 1933–1942
- Kirsch C, Takamiya-Wik M, Schmelzer E, Hahlbrock K, Somssich IE (2000) A novel regulatory element involved in rapid activation of parsley ELI7 gene family members by fungal elicitor or pathogen infection. *Mol Plant Pathol* **1**: 243–251
- Knispel N, Ostrozhenkova E, Schramek N, Huber C, Peña-Rodríguez LM, Bonfill M, Palazón J, Wischmann G, Cusidó RM, Eisenreich W (2013) Biosynthesis of panaxylinol and panaxydol in *Panax ginseng*. *Molecules* **18**: 7686–7698
- Konovalov DA (2014) Polyacetylene compounds of plants of the Asteraceae family. *Pharm Chem J* **48**: 613–631
- Kozlov A, Darriba D, Flouri T, Morel B, Stamatakis A (2018) RAXML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*: btz305
- Lakhssassi N, Zhou Z, Liu S, Colantonio V, AbuGhazaleh A, Meksem K (2017) Characterization of the *FAD2* gene family in soybean reveals the limitations of gel-based TILLING in genes with high copy number. *Front Plant Sci* **8**: 324
- Lee M, Lenman M, Banaś A, Bafor M, Singh S, Schweizer M, Nilsson R, Liljeborg C, Dahlqvist A, Gummeson PO, et al (1998) Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. *Science* **280**: 915–918
- Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, et al (2013) Acyl-lipid metabolism. *The Arabidopsis Book* **11**: e0161, doi:10.1199/tab.0161
- Lopez-Nieves S, Yang Y, Timoneda A, Wang M, Feng T, Smith SA, Brockington SF, Maeda HA (2018) Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. *New Phytol* **217**: 896–908
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol* **207**: 437–453
- Martínez-Rivas JM, Sperling P, Lühs W, Heinz E (2001) Spatial and temporal regulation of three different microsomal oleate desaturase genes (*FAD2*) from normal-type and high-oleic varieties of sunflower (*Helianthus annuus* L.). *Mol Breed* **8**: 159–168
- Meesapyodsuk D, Reed DW, Covello PS, Qiu X (2007) Primary structure, regioselectivity, and evolution of the membrane-bound fatty acid desaturases of *Claviceps purpurea*. *J Biol Chem* **282**: 20191–20199
- Milo R, Last RL (2012) Achieving diversity in the face of constraints: Lessons from metabolism. *Science* **336**: 1663–1667
- Minto RE, Blacklock BJ (2008) Biosynthesis and function of polyacetylenes and allied natural products. *Prog Lipid Res* **47**: 233–306
- Mirarab S, Nguyen N, Guo S, Wang L-S, Kim J, Warnow T (2015) PASTA: Ultra large multiple sequence alignment for nucleotide and amino acid sequences. *J Comput Biol* **22**: 377–386
- Moghe GD, Kruse LH (2018) The study of plant specialized metabolism: Challenges and prospects in the genomics era. *Am J Bot* **105**: 959–962
- Moghe GD, Leong BJ, Hurney SM, Daniel Jones A, Last RL (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* **6**: 1–33
- Moran JH, Mitchell LA, Bradbury JA, Qu W, Zeldin DC, Schnellmann RG, Grant DF (2000) Analysis of the cytotoxic properties of linoleic acid metabolites produced by renal and hepatic P450s. *Toxicol Appl Pharmacol* **168**: 268–279
- Nam J-W, Kappock TJ (2007) Cloning and transcriptional analysis of *Crepis alpina* fatty acid desaturases affecting the biosynthesis of crepenynic acid. *J Exp Bot* **58**: 1421–1432
- Napier JA (2007) The production of unusual fatty acids in transgenic plants. *Annu Rev Plant Biol* **58**: 295–319
- Negri R (2015) Polyacetylenes from terrestrial plants and fungi: Recent phytochemical and biological advances. *Fitoterapia* **106**: 92–109
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274
- Ohlrogge J, Thrower N, Mhaske V, Stymne S, Baxter M, Yang W, Liu J, Shaw K, Shorrosh B, Zhang M, Wilkerson C, Matthäus B (2018) PlantFADB: A resource for exploring hundreds of plant fatty acid structures synthesized by thousands of plants and their phylogenetic relationships. *Plant J* **96**: 1299–1308
- Okuley J, Lightner J, Feldmann K, Yadav N, Lark E, Browse J (1994a) *Arabidopsis FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *Plant Cell* **6**: 147–158
- Okuley J, Lightner J, Feldmann K, Yadav N, Lark E, Browse J (1994b) *Arabidopsis FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *Plant Cell* **6**: 147–158
- Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. *Trends Plant Sci* **5**: 439–445
- Pond SL, Frost SDW, Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490
- Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, Donoghue PCJ (2018) The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr Biol* **28**: 733–745.e2
- Rawat R, Yu X-H, Sweet M, Shanklin J (2012) Conjugated fatty acid synthesis: Residues 111 and 115 influence product partitioning of *Morinda charantia* conjugase. *J Biol Chem* **287**: 16230–16237
- R Core Team (2019) R: A Language and Environment For Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Rognes T (2011) Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC Bioinformatics* **12**: 221
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* **6**: 461–464
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* **32**: 1342–1353
- Sperling P, Ternes P, Zank TK, Heinz E (2003) The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids* **68**: 73–95
- Stull GW, Schori M, Soltis DE, Soltis PS (2018) Character evolution and missing (morphological) data across Asteridae. *Am J Bot* **105**: 470–479

- Talevich E, Invergo BM, Cock PJA, Chapman BA** (2012) Bio.Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics* **13**: 209
- Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH** (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res* **18**: 1944–1954
- Tank DC, Donoghue MJ** (2010) Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and taxa. *Syst Bot* **35**: 425–441
- The Angiosperm Phylogeny Group** (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* **181**: 1–20
- van de Loo FJ, Broun P, Turner S, Somerville C** (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. *Proc Natl Acad Sci USA* **92**: 6743–6747
- Wagih O** (2017) ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics* **33**: 3645–3647
- Weng JK, Philippe RN, Noel JP** (2012) The rise of chemodiversity in plants. *Science* **336**: 1667–1670
- Yang Z** (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, et al** (2015) Dissecting molecular evolution in the highly diverse plant clade caryophyllales using transcriptome sequencing. *Mol Biol Evol* **32**: 2001–2014
- Zhang D, Pirtle IL, Park SJ, Nampaisansuk M, Neogi P, Wanjie SW, Pirtle RM, Chapman KD** (2009) Identification and expression of a new delta-12 fatty acid desaturase (*FAD2-4*) gene in upland cotton and its functional expression in yeast and *Arabidopsis thaliana* plants. *Plant Physiol Biochem* **47**: 462–471